



DISTRICT LEADER'S GUIDE

SELECTING AND USING MULTIPLE MEASURES

TO EVALUATE EDUCATORS



"I feel my principal and district leaders care about what I am doing. They want to help me be a better teacher. I like that."

-TEACHER, BRIGHTON 27J

IMPLEMENTING EVALUATION SYSTEMS:

Learning from Pioneering Districts

In an effort to help districts implement local educator evaluation systems in line with the requirements of SB191, the Colorado Legacy Foundation is providing a suite of resources to district and school leaders. CLF worked with three Colorado districts who recently revised their evaluation process. Although these districts implemented their new systems prior to the passage of SB191, the process that each district went through is informative. We hope that highlighting these districts and the lessons they learned along the way will help other districts leverage their success and avoid re-living their most difficult challenges.

Three case studies provide the foundation for this work. District and school leaders can read the case studies for examples of how different districts have approached similar goals. A series of District Leader's Guides build on the case studies and provide more direct and specific guidance to district leaders as they move forward with implementation. The case studies and guides can be found on the [Colorado Legacy Foundation's](#) website.

Brighton Case Study - Brighton educators and administration agreed that their evaluation system was a "dog and pony show." In 2009 they revised their system, with terrific buy-in from the union, to more meaningfully support teachers.

Eagle Case Study - Eagle has spent nearly a decade developing their evaluation system and aligning it to instruction, assessment, and professional development.

Harrison Case Study - In 2007, Harrison hired a new superintendent who instituted a new evaluation system along with rigorous instructional supports, interim assessments and a pay-for-performance system.



WHO SHOULD USE THIS GUIDE?

This guide has been developed from the lessons learned from Brighton, Eagle and Harrison. District leaders should use it in conjunction with the case studies to prepare the district for a new evaluation system. This guide is developed for district leaders who:

- Have read [Preparing Your District to Implement a New Educator Evaluation System and Communicating Effectively with Stakeholders](#).
- Have established a district evaluation team which is looking for examples, lessons learned and implementation tips from Colorado school districts.
- Are prepared to explore, with the district implementation team, the measures the district will use in evaluating teacher and principal effectiveness.



The desired outcomes and practices of any evaluation system will determine the measures used to assess them.



INTRODUCTION

The desired outcomes and practices of any evaluation system will determine the measures used to assess them. Selecting appropriate measures is a critical component. Measures should yield reliable and valid information on whether quality teaching standards have been met and if there has been growth in student achievement. Similarly, principals should be evaluated on their results as an instructional leader and their schools' student achievement.

SB-191 requires that multiple measures be used to allow for a more comprehensive view of a teacher's effectiveness based on several sources of evidence. Multiple measures contribute to greater confidence in the results of the teacher evaluations, provide a more complete picture of teachers' contributions to student learning, contribute to teachers' professional growth; and give more complete evidence about students' learning growth. In fact, the most recent research suggests that using multiple measures, including multiple observations, student feedback, and measures of student achievement, is a better predictor of an effective teacher than other more traditional measures such as graduate degrees or number of years teaching.

THE POWER OF MULTIPLE MEASURES

The Measures of Effective Teaching (MET) project is a research partnership of academics, teachers, and education organizations committed to investigating better ways to identify and develop effective teaching. Nearly 3,000 teachers opened their classrooms to researchers investigating a number of alternative approaches to identifying effective teaching, including: using systematic classroom observations; surveys collecting confidential student feedback; a new assessment of teachers' pedagogical content knowledge; and different measures of student achievement. [**The most recent report documenting the power of using multiple measures is available online.**](#)
[**This is a brief summary of findings.**](#)

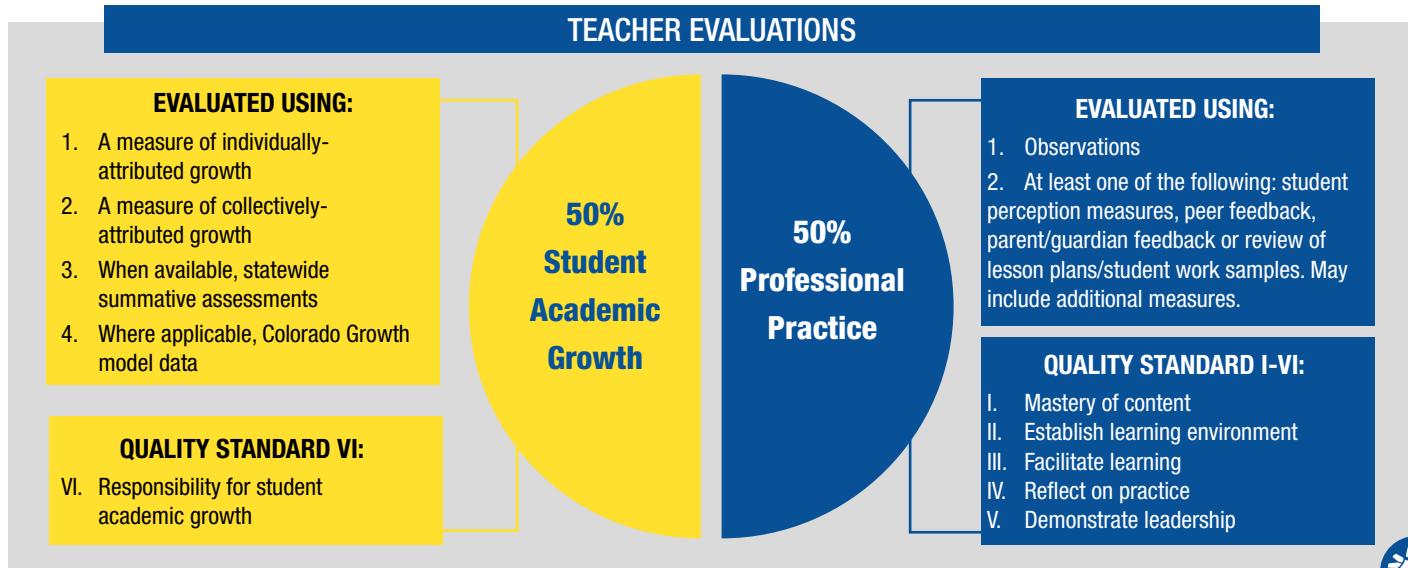
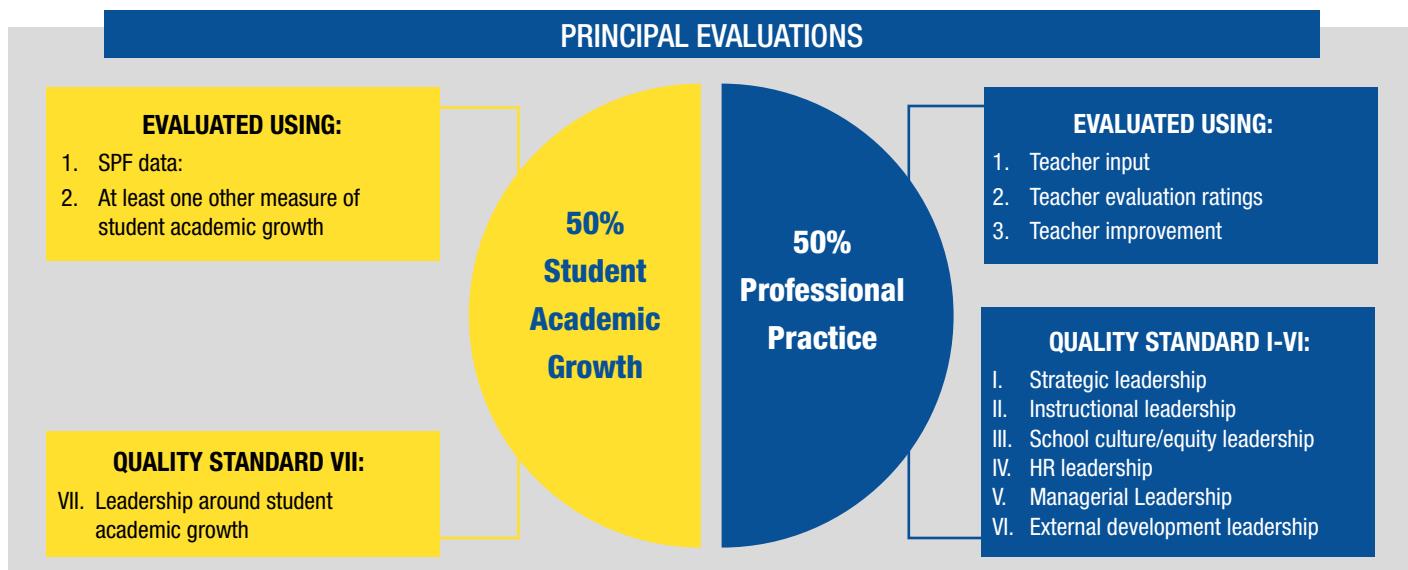
- 1 All five instruments studied in the MET project were positively associated with student achievement gains.
- 2 Reliably characterizing a teacher's practice required averaging scores over multiple observations.
- 3 Combining observations scores with evidence of student achievement gains on state tests and student feedback improved predictive power and reliability.
- 4 Combining observation scores, student feedback, and student achievement gains was better than graduate degrees or years of teaching experience at predicting a teacher's student achievement gains with another group of students on the state tests.
- 5 Combining observation scores, student feedback, and student achievement gains on state tests also was better than graduate degrees or years of teaching experience in identifying teachers whose students performed well on other measures.



GUIDANCE FROM CDE AND THE STATE BOARD OF EDUCATION ON MULTIPLE MEASURES

As of November 2011, the [**Colorado State Board of Education \(SBE\) has approved rules**](#) that will guide the development and implementation of the state model evaluation system. Districts have the option to use the state developed system, or to ensure that their system meets or exceeds criteria specified through the SB 10-191 Rules and Regulations. These Rules will undergo legislative review during the 2012 legislative session to finalize requirements. They are also subject to a continuous improvement process throughout the piloting and implementation of the system.

Fifty percent of a teacher's effectiveness rating will be based on evaluation of their professional practice. The other fifty percent will be based on student academic growth.



GUIDANCE FROM CDE AND THE STATE BOARD OF EDUCATION ON MULTIPLE MEASURES

5.01(E)(6) METHOD FOR EVALUATING TEACHER PROFESSIONAL PRACTICE

With regard to developing multiple measures for evaluating teachers on professional practice, the rules state:

No later than July 2013, a description of the method for evaluating Teachers' Professional Practice, which method shall include data collection for multiple measures on multiple occasions. School Districts and BOCES shall collect Teacher performance data related to Professional Practice using observations and at least one of the following measures: (a) student perception measures (e.g. surveys), where appropriate and feasible, (b) peer feedback, (c) feedback from parents or guardians; or (d) review of Teacher lesson plans or student work samples.

The method for evaluating Teachers' Professional Practice may include additional measures. The Department also shall provide technical guidance, based on research and best practices that emerge from the pilot of the State Model System and the implementation of other local systems during the Pilot Period that School Districts and BOCES may use in developing their own measures of Professional Practice.

In determining how to use the data collected about Teacher performance, whether for written evaluation reports or for informal feedback and identification of appropriate professional development, School Districts and BOCES shall consider the technical quality and rigor of the methods used to collect the data, and the technical quality of the data itself.

With regard to developing multiple measures for evaluating teachers on student growth, the rules state:

Student Academic Growth shall be measured using multiple measures. When compiling these measures to evaluate performance against Teacher Quality Standard VI, School Districts and BOCES shall consider the relative technical quality and rigor of the various measures.

Measures of Student Academic Growth shall include the following:

5.01 (E) (7) (a) A measure of individually-attributed Student Academic Growth, meaning that outcomes on that measure are attributed to an individual licensed person;

5.01 (E) (7) (b) A measure of collectively-attributed Student Academic Growth, whether on a school-wide basis or across grades or subjects, meaning that outcomes on that measure are attributed to at least two licensed personnel (e.g., measures included in the school performance framework, required pursuant to section 22-11-204, C.R.S.);

5.01 (E) (7) (c) When available, Statewide Summative Assessment results; and

5.01 (E) (7) (d) For subjects with annual Statewide Summative Assessment results available in two consecutive grades, results from the Colorado Growth Model.

GUIDANCE FROM CDE AND THE STATE BOARD OF EDUCATION ON MULTIPLE MEASURES

State officials acknowledge that there are a lot of unknowns regarding how multiple measures will be implemented with regard to student growth. However, through the establishment of Content Collaboratives, Colorado educators and state and national experts will establish examples of student growth measures within each content area. During 2012 the primary focus of the work of the Content Collaboratives will be the development of valid, reliable and fair student measurement tools which indicate student academic growth and can be used to evaluate educator effectiveness. These will be piloted in select districts, undergo revision and then be disseminated to all districts. In future years, the focus of common tasks will include the establishment of outstanding instructional lessons which support the academic standards. The products created each year are expected to help schools and districts avoid the need to create education tools by themselves and to build capacity statewide.

Visit [**CDE's website**](#) to learn more and track the work of the Content Collaboratives.

MATCHING INTENDED USE OF DATA TO STUDENT OUTCOME MEASURES

Possible uses of the evaluation system:

- Rating individual performance
- Informing personnel decisions
- Providing educators relevant and useful feedback which can be used to improve instruction.
- Urgency of improving student outcomes.
- Fairness of a system to individual educators.
- Need for a system to be simple enough to be implemented well in a wide variety of districts across the state, and to account for differing district priorities.
- Opportunity to develop measures of student growth and achievement in all areas of Colorado Academic Standards (fairness to teachers in those areas, emphasize importance of these content areas, etc)
- Opportunity to develop new models of measuring student growth and achievement that can inform the assessment of currently tested subjects and grades.
- Enable district- and school-level choice regarding which measures of student growth are most appropriate to meet local needs and objectives.
- Need to ensure that all methods of measuring student growth satisfy some minimum level of technical rigor regarding the calculation of student growth and combining multiple measures.
- Highly varied and often limited local resources available to develop, administer and analyze measures of student growth.
- Impact of small class sizes in calculating student growth.
- Ensuring that determinations of an individual educator's performance are based upon credible evidence that clearly supports the performance rating.

*State Council on Educator Effectiveness Recommendations
Recommendation 14 p. 69-70.*



IMPORTANT CONSIDERATIONS FOR DISTRICTS

Measures should be rigorous, measure student growth and be comparable across classrooms and within a district. All measures selected should be valid and reliable for their intended purposes. That means they must accurately and fairly measure what standards the student is supposed to learn, whether the student has learned it and how results can be attributed to individual teachers.

Other states that are combining measures of professional practice with student academic growth measures are finding that many teachers are scoring high on the elements of professional practice but low on academic growth. They believe this is in large part due to lack of training or the need for additional training of evaluators observing professional practice. (The issue of selecting and training evaluators is addressed in a separate CLF guide.)

The table below, from *A Practical Guide to Designing Comprehensive Teacher Evaluation Systems* suggests how purposes of evaluation systems can be matched with a wide variety of measures – many of which districts might already have on hand. “It reviews potential teacher evaluation goals and identifies the measurement types that are most appropriate to meet those goals.”

Purpose of Evaluation of Teacher Effectiveness Value	Value-Added	Classroom Observation	Analysis of Artifacts	Portfolios	Teacher Self Reports	Student Ratings	Other Reports
Find out whether grade-level or instructional teams are meeting specific achievement goals.	x						
Determine whether a teacher's students are meeting achievement growth expectations.	x	x					
Gather information in order to provide new teachers with guidance related to identified strengths and shortcomings.	x	x	x	x			x
Examine the effectiveness of teachers in lower elementary grades for which no test scores from previous years are available to predict student achievement (required for value-added models).	x	x	x	x			x
Examine the effectiveness of teachers in nonacademic subjects (e.g., art, music, and physical education).	x		x		x	x	x
Determine whether a new teacher is meeting performance expectations in the classroom.	x	x	x		x	x	x
Determine the types of assistance and support a struggling teacher may need.	x	x		x	x		
Gather information to determine what professional development opportunities are needed for individual teachers, instructional teams, grade-level teams, etc.	x	x		x			x
Gather evidence for making contract renewal and tenure decisions.	x	x					x
Determine whether a teacher's performance qualifies him or her for additional compensation or incentive pay (rewards).	x	x					
Gather information on a teacher's ability to work collaboratively with colleagues to evaluate needs of and determine appropriate instruction for at-risk or struggling students.				x	x		x
Establish whether a teacher is effectively communicating with parents/guardians.				x			x
Determine how students and parents perceive a teacher's instructional efforts.						x	
Determine who would qualify to become a mentor, coach, or teacher leader.	x	x	x	x			x



LEARNING FROM PIONEERING DISTRICTS

GUIDING QUESTIONS

As district leaders and the district implementation team consider how to go about selecting multiple measures, consider these guiding questions.

- Do the selected measures provide data to inform progress on the evaluation system goals?
- Does the measure match the purpose of the evaluation?
- Does the measure provide valid and reliable data to make high-stakes decisions such as placing a teacher in a performance standard?
- Does the measure have research on its
 - Ability to measure student progress?
 - Demonstrated impact on student achievement?
 - Demonstrated impact on teacher practice?
- What processes are in place to ensure the technical rigor of the measure?
- Is the measure an accurate and fair indicator of what a student is supposed to learn?
- Is the measure an accurate and fair indicator of teacher practice?
- Do student populations and the context of teaching need to be differentiated to provide reliable and valid data?
- Can the measure be implemented with limited human and resource capacity? What resource capacity is necessary?
- Can the measure of student growth be attributed accurately to multiple teachers?
- Are key data collected that provide information on the district's status and progress toward reaching its goals and priorities?

LEARNING FROM PIONEERING DISTRICTS

Although Brighton, Eagle and Harrison undertook reform of their evaluation systems prior to the passage of SB191, the process by which they developed the system and rolled it out district wide yielded many lessons learned and implementation tips. District evaluation teams are encouraged to [read the case studies and use the case study highlights](#) to further inform their discussions.

The following sections include examples and lessons learned from the district case studies.

INSTRUCTIONALLY FOCUSED ACCOUNTABILITY – EAGLE COUNTY SCHOOLS

Eagle School District measures teacher performance and student learning in three ways:

1. Teacher skills, knowledge and responsibility based on a teacher evaluation rubric are focused on four domains:
 - Planning, instruction and assessment (2 indicators)
 - Instruction (9 indicators)
 - Learning environment (5 indicators)
 - Professional responsibilities (6 indicators)
2. School-wide achievement is based on Colorado Student Assessment Program (CSAP) tests and the ACT.
3. Individual student achievement is based on the Northwest Evaluation Association's (NWEA) Measures of Academic Progress (MAP) benchmark assessment system. A statistical model was used to create growth scores for individual students.

During the 2011-2012 school year Eagle County Schools will be using district-developed "rigorous curriculum design" units in mathematics, literacy, science and social studies. The aligned curriculum components provide clear learning outcomes with matching assessments along with engaging learning experiences and instructional strategies. The new assessments will replace the NWEA MAP tests as individual student measures.



IMPLEMENTATION TIPS AND LESSONS LEARNED

Pioneering districts suggested the following implementation tips and lessons learned:

- *Use the CSAP growth model.* The state assessment is useful in providing individual student growth scores that could be compared with district and state results.
- *Create performance measures in non-tested subjects.* Harrison Two developed creative performance measures in areas such as art and physical education. Rubrics were used to score student work.
- *Teachers want credible measures to assess student achievement. Creating common assessments can help achieve this.* A trend in the pioneering districts was to develop their own common assessment and curriculum based measures. Two districts engaged teachers from across the district in writing common assessment items which were later vetted and included as student measures.
- *Teachers need to know what quality standards look like in practice.* When teachers are judged on standards, especially those related to instruction, they need to know what performance looks like at each level of the rubric. Videos, modeling practices and coaching are useful strategies. Two districts had well defined rubrics with clear descriptions on which to assess teacher performance in all types of observations – from informal spot observations to annual summative evaluations.
- *Get results back quickly.* The sooner teachers have feedback, the sooner they can use the data to improve teaching and learning.
- *Teachers valued feedback on their instructional practices.* Some teachers put more credibility on spot observations because they were more representative of their daily teaching practice than formal evaluations where they practiced for an announced evaluation. Virtually all teachers wanted to improve their instructional practice and coaching on their practices helped them improve.

In summary, selecting appropriate and multiple measures is a critical component of the design process. Measures should yield reliable and valid information on whether quality teaching standards have been met and if there has been growth in student achievement. Measures should also be selected to assess the goals of the evaluation system.

ACKNOWLEDGMENTS

Author: Jane Armstrong, JM Armstrong and Associates

Research Support and Editing: Heather Chikoore, Colorado Legacy Foundation

Research Support for the Case Studies: Ulcca Hansen, Colorado Legacy Foundation and Kristen Davidson,
University of Colorado at Boulder

- 1 J. Herman, M. Heritage and P. Goldschmidt. Measuring Teachers Contributions to Student Learning Growth for Nontested Grades and Subjects. (Washington, DC: National Comprehensive Center for Teacher Quality). 2011.
- 2 Laura Goe, Lynn Holdhelde and Tricia Miller. *A Practical Guide to Designing Comprehensive Teacher Evaluation Systems.* (Washington, DC: National Comprehensive Center for Teacher Quality). May 2011.
- 3 Laura Goe, Lynn Holdhelde and Tricia Miller. *A Practical Guide to Designing Comprehensive Teacher Evaluation Systems.* (Washington, DC: National Comprehensive Center for Teacher Quality). May 2011.